Research Article

pISSN 1225-6161 • eISSN 2287-9307
Korean Journal of Remote Sensing 2024, 40(6-3), 1505-1521
https://doi.org/10.7780/kjrs.2024.40.6.3.11

Korean Journal of Remote Sensing
KJRS

SAMBA: Synthetic Data-Augmented Mamba-Based Change Detection Algorithm Using KOMPSAT-3A Imagery

Rogelio Ruzcko Tobias[1†], Sejeong Bae[2†], Hwanhee Cho[3], Jungho Im[4*]

[1]PhD Student, Artificial Intelligence Graduate School, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[2]Combined MS/PhD Student, Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
[3]Undergraduate Student, Department of Convergence and Fusion System Engineering, Kyungpook National University, Sangju, Republic of Korea
[4]Professor, Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

Abstract: Change detection is essential for applications such as urban planning, environmental monitoring, and disaster response. Despite advancements in high-resolution satellite imagery, accurate change detection remains challenging due to increased landscape heterogeneity and variable atmospheric conditions. The Mamba model, an efficient state-space model-based architecture, has shown promise in capturing spatiotemporal relationships in high-resolution datasets, addressing the limitations of traditional methods that struggle with the diverse appearances of urban structures. This research investigates applying Mamba to multitemporal Korea Multi-Purpose Satellite (KOMPSAT) imagery, using both real and synthetic data from SyntheWorld, a dataset developed to simulate various change scenarios. This study introduces a synthetic data-augmented mamba-based change detection algorithm (SAMBA), designed to detect structural changes in urban environments using KOMPSAT-3A satellite imagery. The main objectives are to evaluate the Mamba binary change detection (MambaBCD) model's ability to detect building changes in KOMPSAT-3A images and assess the impact of synthetic data augmentation on performance. Experimental results with MambaBCD-Small and MambaBCD-Tiny models indicate that synthetic data incorporation improves generalization in complex settings, achieving high performance across multiple data and model configurations. Notably, the MambaBCD-Tiny model, with or without synthetic augmentation, outperformed the larger-parameter MambaBCD-Small model, demonstrating enhanced sensitivity in detecting satellite image changes. Performance evaluation metrics yielded an overall accuracy of 99.73%, precision of 98.34%, recall of 96.54%, F1-score of 97.43%, intersection over union of 95.00%, and Kappa coefficient of 97.29%. These metrics were similarly used to test the SAMBA algorithm's generalization on benchmark change detection datasets, showcasing its potential as a robust tool for high-resolution image change detection.

Keywords: KOMPSAT, Remote sensing, Change detection, Artificial intelligence, Computer vision, Mamba

Received: November 19, 2024
Revised: December 2, 2024
Accepted: December 3, 2024
Published: December 31, 2024

Corresponding author:
Jungho-Im
E-mail: ersgis@unist.ac.kr

1. Introduction

Accurate detection of changes in spatial information of buildings is critical for monitoring man-made landcover changes, urban planning, and disaster response assessment (Cheng et al., 2024). The launch of various Earth observation satellites equipped with very high-resolution (VHR) sensors have increased accessibility to frequent high-resolution images in areas of interest. Through

† These authors contributed equally to this work.

the fusion of VHR imagery—which can provide detailed structural and textural information about surface features—and rapidly advancing image processing technologies, change detection research in urban areas is considered one of the most important subtasks using optical satellite imagery (Hussain et al., 2013; Park and Song, 2023).

Detecting changes in buildings and urban artificial structures caused by human beings has been one of the topics of greatest interest, and various studies have been conducted to detect building changes using VHR imagery (Song et al., 2020; Lu et al., 2024). However, the increase in spatial resolution (Chen et al., 2024a) brings about increased heterogeneity within the same landcover feature, leading to limitations in information extraction relying on manual visual interpretation. As the resolution of satellite imagery increases, the sensitivity to uncertainties in preprocessing steps such as radiometric correction, geometric correction, and image registration also increases (Im et al., 2008). Furthermore, building objects in satellite images have varying shapes and appearances, and the same building object at different times may have distinct colors due to illumination variations and appearance alterations (Chen et al., 2021). Therefore, developing building change detection frameworks in urban areas that can withstand complex and irrelevant discrepancies is a challenging task.

Nowadays, owing to their powerful discriminative abilities, deep learning-based change detection algorithms that can effectively process multiple datasets have been introduced in the fields of remote sensing and computer vision: fully convolutional network (FCN) (Lee et al., 2021; Chen et al., 2024b), vision transformer (ViT) (Dosovitskiy et al., 2020), and state-space model (SSM)-based algorithms (Gu and Dao, 2024). FCN models have been actively used for change detection in the domain of satellite imagery (Daudt et al., 2018; Song et al., 2020). They overcome the limitation of convolutional neural networks (CNNs) which learn relationships around specific pixels through kernels but cannot process temporal information within their structures, by allowing end-to-end learning for pixel-wise detection. Additionally, ViT-based models have demonstrated remarkable performance in capturing global contextual information by leveraging self-attention mechanisms. This ability enables ViT to handle complex and diverse multitemporal scenes in images with different spatial-temporal resolutions better than FCNs (Chen et al., 2021). At the same time, the self-attention mechanism in Transformers poses challenges in terms of speed and memory usage when dealing with long-range visual dependencies, especially VHR satellite imagery (Zhu et al., 2024).

SSM-based models have emerged as a viable alternative to the Transformer architecture (Vaswani et al., 2017), exhibiting state-of-the-art performance in analyzing long-sequence data. Specifically, the Mamba architecture, an enhancement of the structured state-space sequence (S4) model, has been applied to high-resolution public datasets (e.g., SYSU-CD, LEVIR-CD, and WHU-CD) and has shown promising results in change detection tasks (Chen et al., 2024a; Paranjape et al., 2024). Mamba employs a selection mechanism that enables the model to choose relevant information in an input-dependent manner, achieving linear scaling with sequence length. Unlike existing change detection methods using VHR satellite imagery—which often face challenges in computational efficiency and capturing long-range dependencies—the Mamba architecture uniquely combines its input-dependent selection mechanism with hardware-aware implementations (Chen et al., 2024a). This combination allows it to effectively capture global spatial context and spatiotemporal relationships in VHR satellite imagery while maintaining computational efficiency.

However, in the real-world application, only a small portion changes within the large imaging area of satellite data, and there is higher uncertainty compared to public datasets used in computer vision change detection research. Moreover, data from satellites that provide VHR images are difficult to acquire and, compared to aerial images, are more affected by atmospheric conditions or weather at the time of image capture, which can lead to differences in local radiometric characteristics. Synthetic data, as an additional dataset, can effectively deal with these variations by including various application scenarios within the training process. In the field of computer vision, numerous high-quality synthetic datasets have emerged, serving tasks such as building change detection, landcover classification, and segmentation (Song et al., 2024). Song et al. (2024) proposed SyntheWorld, a specific synthetic dataset for building change detection and landcover classification. By incorporating the synthetic dataset as additional input features added to VHR images during the training process, they demonstrated increased performance in change detection tasks using deep CNN and ViT-based models (Chen et al., 2021; Peng et al., 2019; Zhang et al., 2022).

In this study, we aimed to apply the Mamba model to the change detection task using images from the current operational VHR satellite, the Korea Multi-Purpose Satellite-3A (KOMPSAT-3A). Since there are limitations in acquiring KOMPSAT images,

 https://doi.org/10.7780/kjrs.2024.40.6.3.11

Table 1. Detailed information (the acquisition dates, number of image pairs, image sizes, spatial resolutions, and file formats) of change detection datasets used in this study

Dataset	Date	Image pairs	Image size	Spatial resolution	Format
KOMPSAT	2015.10.28 2023.10.06	1	3,200 x 3,200	0.55 m	TIFF
SYSU-CD	2007 2014	20,000	256 x 256	0.5 m	PNG
LEVIR-CD+	2002 2018	985	1,024 x 1,024	0.3 m	PNG
WHU-CD	2012 2016	1	32,207 x 1,535	0.2 m	PNG
SyntheWorld	- -	40,000*	1,024 x 1,024	0.3 m	PNG

we incorporated SyntheWorld as an additional input dataset. The specific objectives of this research were to: (1) evaluate whether the Mamba model can detect building changes in multitemporal VHR images from KOMPSAT and (2) assess whether incorporating synthetic datasets into the Mamba can improve performance. This study demonstrated the effectiveness of the current state-of-the-art (SOTA) deep learning architecture, Mamba, for binary change detection (BCD) of buildings in real satellite imagery. In addition, this study provided insights into performance improvements of deep learning BCD models when utilizing synthetic datasets.

2. Data

2.1. KOMPSAT

Operationally, KOMPSAT-3A, the sister satellite of KOMPSAT-3, uses optical high-resolution sensors to provide information applicable to geographical information systems, environmental monitoring, and agricultural purposes (Table 1). Researchers have actively utilized KOMPSAT-3 and KOMPSAT-3A VHR imagery, especially in the fields of change detection and extracting building information in urban areas (Han et al., 2017; Lee et al., 2024; Song et al., 2020). KOMPSAT-3A operates at an altitude of 528 km and has a ground sample distance of 0.55 m (Jeon et al., 2016; Kim et al., 2020). In this study, we used a single pair of images acquired on October 28th, 2015, and October 6th, 2023, covering the same area in Sejong City, South Korea, with a grid size of 1.76 km as seen in Fig. 1. The images were pansharpened RGB images with geometric corrections applied, provided by the 2024 Satellite Information Utilization Competition.

The change detection reference data for this study were provided by the Korea Aerospace Research Institute (KARI) and

Fig. 1. Bitemporal Korea Multi-Purpose Satellite-3A (KOMPSAT-3A) images of Sejong City, South Korea. (T1) refers to the pre-event image taken on October 28th, 2015, (T2) refers to the post-event image taken on October 6th, 2023, and (GT) is the ground truth binary change map for the two given images.

derived from a single pair of KOMPSAT-3A images (Fig. 1). These reference datasets consist of binary classification maps of building changes, each possessing a spatial resolution of 0.55 meters, which aligns with the resolution of the KOMPSAT-3A satellite imagery utilized in this research. The criteria for change detection were established based on four specific types of building changes. Construction of new buildings refers to the process of erecting a new structure on an existing vacant lot, resulting in a detected change mask corresponding to the new building form. Destruction of existing buildings and subsequent construction of new buildings on the same site involves demolishing an existing building, converting the area into a vacant lot, and then constructing a new building, for which a changes mask reflecting the original building form is generated. Reconstruction of buildings pertains to the renovation or rebuilding of an existing building in a different form, necessitating a change mask corresponding to the original building form to capture these alterations. Buildings under construction describe the completion process of an ongoing construction project, resulting in a detected changes mask for the completed building form.

2.2. SyntheWorld

SyntheWorld (Song et al., 2024) is a large-scale synthetic dataset developed for remote sensing VHR image processing tasks, focusing on building change detection and landcover classification, consisting of a total of 40,000 pairs of multitemporal images (Table 1). The dataset comprises image patches of size 1,024 × 1,024 pixels, with resolutions ranging from 0.3 m to 1.0 m, specifically grouped into two ranges: 0.3 m to 0.6 m and 0.6 m to 1.0 m. It is procedurally generated using Blender and AI-generated content techniques and includes change information across various urban and rural environments. The dataset encompasses a wide range of urban styles, buildings, trees, terrains, and more, effectively reflecting complex urban structures and changes. Images have been created for various situations to adequately account for sensor angle distortions and the diversity of illumination angles that can occur in actual RGB images. Moreover, SyntheWorld provides BCD maps of building and landcover annotations for eight classes. In this study, not all the 40,000 images of this dataset were used. We only utilized 145 pairs of synthetic data with small off-nadir (SN) and 145 pairs of synthetic data with big off-nadir (BN) images representing pre-event and post-event following the recommended configuration of 7:1 real-to-synthetic data ratio from (Song et al., 2024).

2.3. Benchmark Datasets

Three benchmark datasets were additionally used in this study. First, the SYSU-CD (Shi et al., 2022) dataset is a public BCD dataset consisting of 20,000 pairs of VHR aerial images taken in Hong Kong (Table 1). Based on imagery with a resolution of 0.5 m captured between 2007 and 2014, it primarily includes various change information in urban areas. The dataset comprises a total of 20,000 pairs of multitemporal image patches and their corresponding binary patches, which effectively reflect complex urban structures and high-rise building changes.

LEVIR-CD is a public, large-scale building change detection dataset (Chen and Shi, 2020). It provides remotely sensed VHR images and corresponding BCD maps, consisting of 637 pairs with a ground sample distance of 0.5 m and image patches of size 1,024 × 1,024 pixels (Table 1). The bitemporal images in LEVIR-CD are from 20 different regions and contain variations due to seasonal changes and illumination differences.

WHU-CD (Liu and Ji, 2020) is a public building change detection dataset containing a pair of VHR aerial images from 2012 and 2016 with a ground sample distance of 0.075 m and a BCD map (Table 1). The images cover an area of 20.5 km^2, featuring building changes in regions rebuilt after the 2011 earthquake. The dataset is provided with images that have been georeferenced using 30 control points, achieving an accuracy of 1.6 pixels.

3. Methods

To perform BCD based on VHR optical satellite data, this study utilized the Mamba algorithm, which applies an SSM-based model to computer vision. We aimed to apply the Mamba model to change detection tasks using images from the KOMPSAT-3A. Due to limitations in acquiring sufficient KOMPSAT-3A images, we incorporated SyntheWorld as an additional synthetic input dataset.

We constructed the input variables for the deep learning model used in change detection by utilizing a pair of pansharpened RGB KOMPSAT-3A images. The target variable was a BCD map of buildings, which was manually created through visual inspection of the respective areas. To investigate whether synthetic data can capture characteristics beyond actual satellite images when training data for deep learning models are insufficient, we compared the performance of BCD across various combinations of input variables and model parameters. Due to the difficulty in acquiring additional VHR optical satellite images, we validated

https://doi.org/10.7780/kjrs.2024.40.6.3.11

the deep learning model by dividing the regions of the KOMPSAT-3A image pair into patches and splitting them into training and validation samples in a 6:4 ratio. Additionally, we evaluated the model's performance using benchmark VHR datasets beyond KOMPSAT-3A to ensure its robustness across different data sources.

3.1. Data Preparation and Scheme Configuration

The preparation of training and validation samples for the deep learning model involved patching a pair of KOMPSAT-3A images. To augment the training data, the 3,200 × 3,200-pixel dataset images were divided into overlapping 1,024 × 1,024-pixel masks, resulting in a total of 1,024 patches. These patches were then combined with SyntheWorld data in a 7:1 ratio to establish three distinct input variable schemes (KOMPSAT, w/ SyntheWorld (SN), and w/ SyntheWorld (BN)). SyntheWorld samples highlight the variability introduced by off-nadir angles from the post-event images, essential for simulating real-world conditions in training data (Fig. 2).

The first scheme (KOMPSAT) utilized only the 1,024 KOMPSAT-3A patches, which were subsequently divided into 614 training samples and 410 validation samples. The second (w/ SyntheWorld (SN)) and third (w/ SyntheWorld (BN)) schemes incorporated SyntheWorld's small off-nadir (SN) and big off-nadir (BN) images, respectively, alongside the KOMPSAT-3A patches. Specifically, each of these schemes combined 1,024 KOMPSAT-3A patches with 145 SyntheWorld images, which were then split into training and validation sets in a 6:4 ratio. Each input variable combination was employed to train and validate the deep learning models, allowing for the evaluation of model performance when synthetic datasets were included.

3.2. Model Architecture

The SSM-based models and the Mamba architecture (Gu and Dao, 2024) are inspired by linear time-invariant systems—mathematical frameworks that process sequences or signals over time. In these systems, an input sequence $x(t)$ is transformed into an output $y(t)$ through an internal hidden state $h(t)$. This transformation relies on specific parameters that define how the system evolves.

Integrating continuous systems directly into deep learning algorithms poses significant challenges. To address this, the S4 model offers a discrete version of these systems, making them more compatible with digital computation. It introduces a time scale parameter to convert the continuous parameters into discrete ones, often using a method called zero-order hold (ZOH).

Once discretized, the system updates its hidden state at each

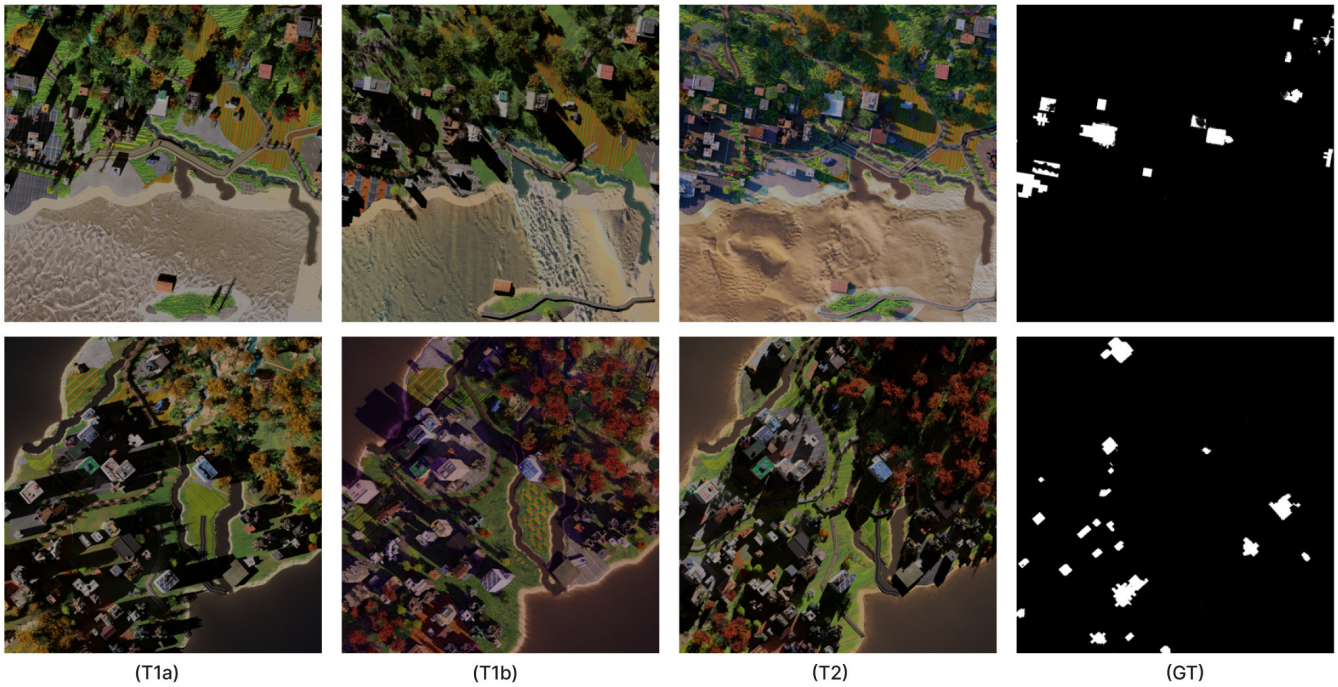

Fig. 2. Sample bitemporal SyntheWorld images from the work of Song et al. (2024). (T1a) refers to the pre-event image with small off-nadir (SN), (T1b) refers to the pre-event image t with big off-nadir (BN), (T2) refers to the post-event image, and (GT) is the ground truth binary change map.

time step based on the previous state and current input, with the output calculated via convolution operations. This method is commonly used in deep learning, applying filters across input sequences to streamline processing. Expanding on the S4 model, the Mamba architecture introduces a selection mechanism to filter out irrelevant information and emphasize critical inputs, improving efficiency and focus. Moreover, Mamba includes a hardware-aware algorithm that optimizes memory usage and computational speed on GPUs, allowing it to avoid storing extensive states and enhancing its overall efficiency.

Building change detection is a primary and well-researched task in change detection, focused on identifying areas of change between two time periods. Building change detection can be divided into two approaches based on the category of interest: Category-Agnostic Change Detection, which detects general land-cover changes without focusing on specific categories, and Single-Category Change Detection, which targets changes within a particular category, such as buildings or forests. Given a training dataset represented as:

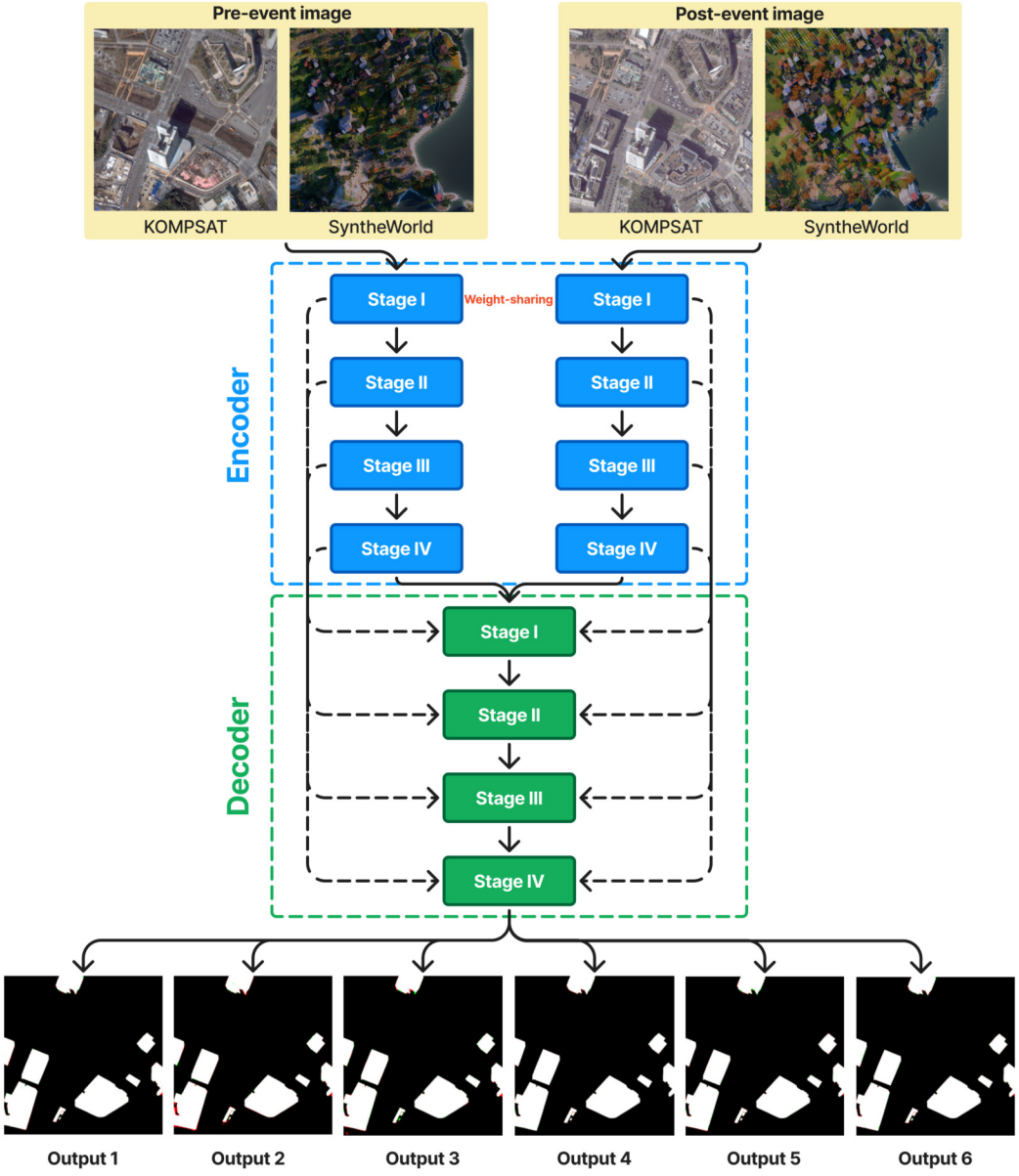

Fig. 3. Architecture of the proposed SAMBA model, illustrating the integration of spatiotemporal state space (STSS) blocks and fusion modules to capture and refine multitemporal relationships for accurate binary change map generation. The specific structure of the encoder and decoder is shown in Figs. 5 and 6.

https://doi.org/10.7780/kjrs.2024.40.6.3.11

$$D_{train}^{bcd} = \{(X_i^{T1}, X_i^{T2}, Y_i^{bcd})\}_{i=1}^{N_{train}^{bcd}} \quad (1)$$

where $X_i^{T1}, X_i^{T2} \in R^{H\times W\times C}$) are in the i-th pair of multitemporal images captured at times T1 and T2, respectively. $Y_i^{bcd} \in \{0, 1\}^{H\times W}$ is the corresponding binary change map label, indicating "change" (1) or "no change" (0) at each pixel location.

The objective of BCD is to train a change detection model F_θ^{bcd} using training data D_{train}^{bcd}. The trained model aims to predict accurate binary change maps for new image pairs, effectively highlighting areas of change. Mathematically, for new pair of images (X^{T1}, X^{T2}), the model predicts:

$$\hat{Y}^{bcd} = F_\theta^{bcd}(X^{T1}, X^{T2}) \quad (2)$$

where $\hat{Y}_i^{bcd} \in \{0, 1\}^{H\times W}$ is the predicted binary change map. The goal is for $\hat{Y}^{bcd}$ to closely match the true change map Y^{bcd}, accurately reflecting the change/no-change information in new datasets.

Based on the common features and requirements of the change detection tasks, we have adopted a network framework called Mamba Binary Change Detection (MambaBCD), derived from the Mamba architecture (Gu and Dao, 2024). Fig. 3 illustrates the architecture of the proposed SAMBA model. The encoder is a weight-sharing Siamese network built upon the Visual State Space Model (VMamba) architecture (Liu et al., 2024). Benefiting from the Mamba architecture and an efficient 2D cross-scan mechanism as shown in Fig. 4, VMamba effectively extracts robust and representative features from the input images for downstream tasks.

MambaBCD includes a change detector designed to learn spatiotemporal relationships from the features extracted by the encoder. The specific structures of the encoder and decoder are

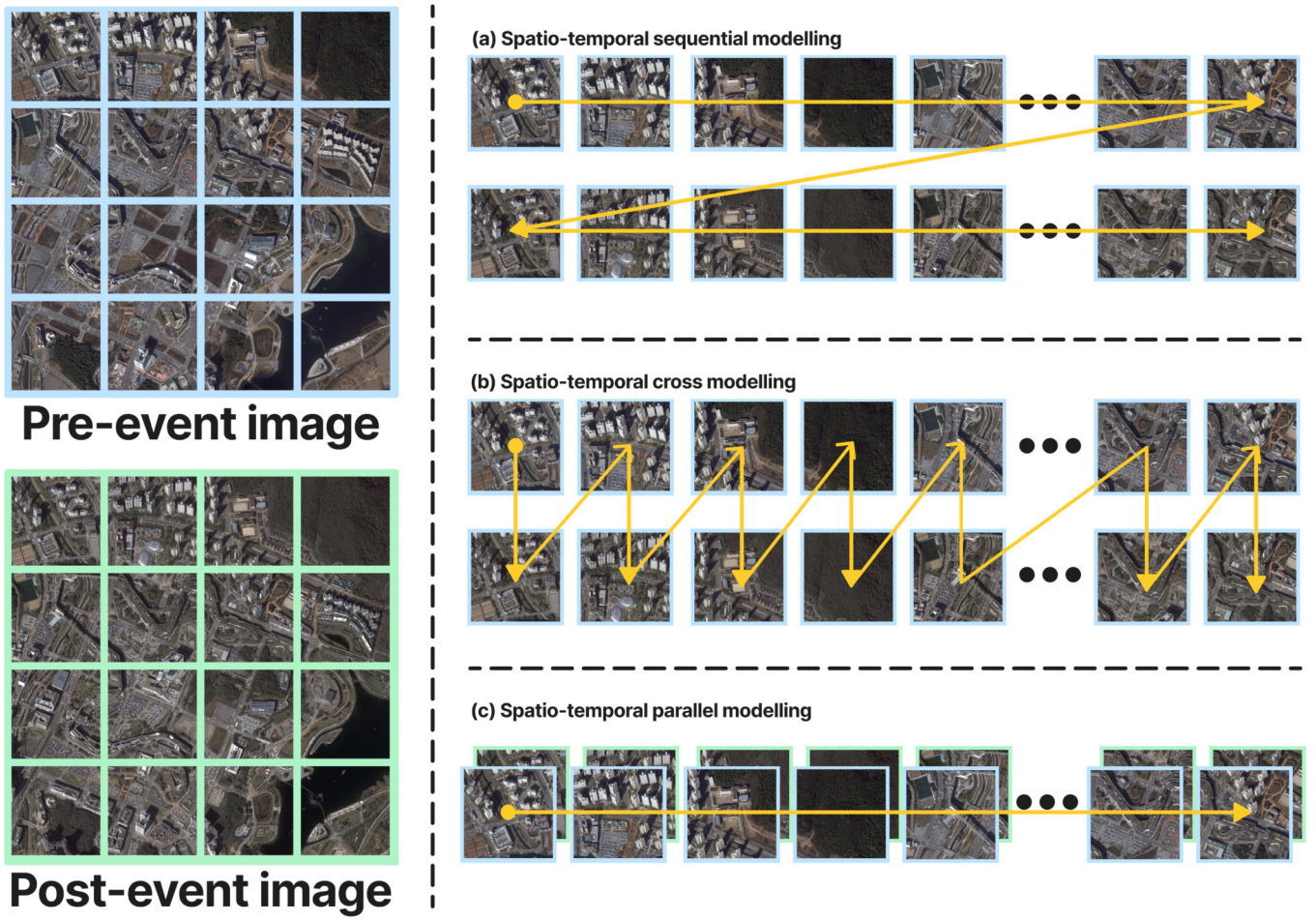

Fig. 4. Illustration of three mechanisms for learning spatiotemporal relationships proposed by Liu et al. (2024) to capture global contextual information. (a) Spatiotemporal sequential modeling. (b) Spatiotemporal cross modeling. (c) Spatiotemporal parallel modeling. Spatiotemporal State Space (STSS) block models the relationships between multitemporal features. The three spatiotemporal relationships modeled by the Visual State Space (VSS) blocks within the STSS block highlight their roles in extracting and learning temporal, spatial, and combined spatiotemporal dependencies from multitemporal features.

detailed in Figs. 5 and 6 (Chen et al., 2024a). In this section, we focus on the inputs, outputs, and internal information flow of the MambaBCD network architecture.

MambaBCD is specifically designed for the Binary Change Detection task. Initially the Siamese encoder network, denoted as F^{θ}_{enc}, this network extracts multi-level features from the input multi-temporal images:

$$\{F^{T1}_{ij}\}^{4}_{j=1} = F^{\theta}_{enc}(X^{T1}_{i}), \{F^{T2}_{ij}\}^{4}_{j=1} = F^{\theta}_{enc}(X^{T2}_{i}) \quad (3)$$

where θ represents the set of parameters (weights) that the

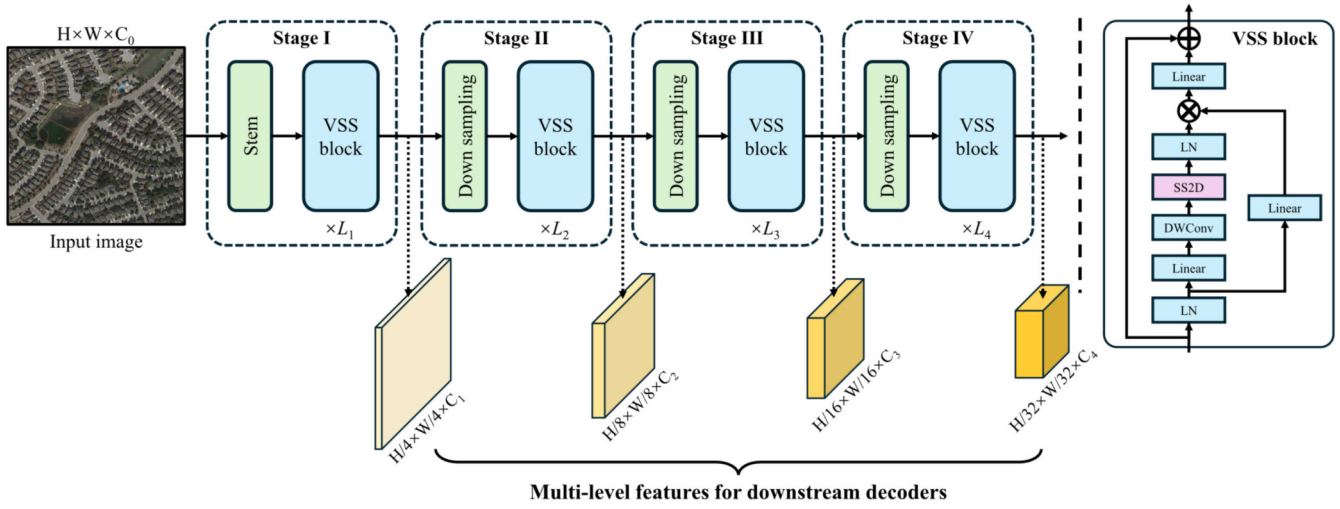

Fig. 5. The encoder network structure is depicted across four stages. Each stage downsamples the input, models spatial contextual information through multiple Visual State Space (VSS) blocks, and outputs features for pre-event and post-event images. The VSS block includes a linear embedding layer, splitting the input into two streams. One stream passes through a 3 × 3 depthwise convolution (DWConv) and a Sigmoid Linear Unit (Silu) activation before entering the Structured State Space for 2D data (SS2D) module, which integrates the Structured State Space model (S6) with a cross-scan mechanism. The outputs are combined after layer normalization and Silu activation, producing the final block output. These features are then passed to decoders for task-specific applications.

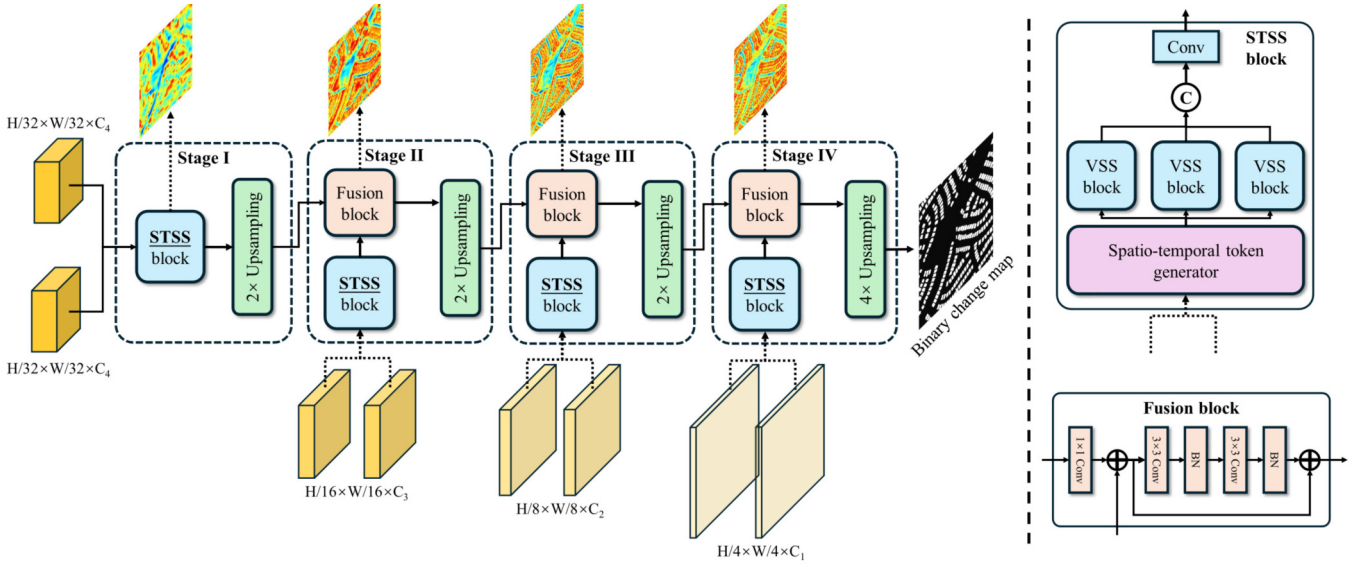

Fig. 6. Structure of the change decoder in the proposed SAMBA model, which leverages three spatiotemporal learning mechanisms based on the work of Chen et al. (2024) to capture relationships within multitemporal features across four stages, producing accurate binary change maps. At each stage, the Spatiotemporal State Space (STSS) block models feature relationships, rearranges input features through a spatiotemporal token generator, and processes them with three Visual State Space (VSS) blocks to learn specific spatiotemporal dependencies as detailed in Fig. 2. The STSS block output is combined with previous stage feature maps in a fusion module, where low- and high-level feature maps are aligned via convolution and summed. The resulting feature map is refined through a residual layer and unsampled before progressing to the next stage.

https://doi.org/10.7780/kjrs.2024.40.6.3.11

network learns during training. X_i^{T1} and X_i^{T2} represent the multi-temporal input images for two different time points, T1 and T2, respectively. The results are $\{F_{ij}^{T1}\}_{j=1}^{4}$ and $\{\{F_{ij}^{T2}\}_{j=1}^{4}$ which denotes the set of features extracted from the inputs, with j representing the four different levels of features. These multi-level features are then fed into a tailored change decoder F_{cdec}^{θ}. Utilizing the Mamba architecture, the change decoder fully learns the spatiotemporal relationships from the multi-level features through three different mechanisms, progressively producing an accurate BCD result formulated as:

$$P_i^{bcd} = F_{cdec}^{\theta}(\{F_{ij}^{T1}\}_{j=1}^{4}, \{F_{ij}^{T2}\}_{j=1}^{4}) \quad (4)$$

where P_i^{bcd} represents the BCD probability map. The binary change map Y_i^{bcd} is then obtained by taking the class with the highest probability for each pixel given by:

$$Y_i^{bcd} = argmax_c\, P_i^{bcd} \quad (5)$$

where the $argmax_c$ function selects the class with the highest probability for each pixel in P_i^{bcd}. In BCD, this often means classifying each pixel as either "change" or "no change," based on which class has the highest probability.

The implementation of the Mamba change detection algorithm closely follows the original implementation in this repository (Chen et al., 2024a) with modifications to work on our current environment, hardware availability, and the inclusion of synthetic data during model training. Depending on the size and depth of the encoder network, these architectures are available in MambaBCD-Tiny, MambaBCD-Small, and MambaBCD-Base versions (Liu et al., 2024). However, due to some limitations in computing power, we only conducted experiments on the MambaBCD-Tiny and MambaBCD-Small configurations. As detailed in Table 2, the main differences among them are the number of Visual State Space (VSS) blocks within each stage and the number of channels in the feature maps.

Table 2. Comparison of the MambaBCD-Tiny and MambaBCD-Small encoder architectures, detailing the number of Visual State Space (VSS) blocks (L) and feature channels (C) at each stage. The configurations highlight the structural differences between the two models in terms of depth and complexity

Stage	Tiny	Small
I	$L_1 = 2$ $C_1 = 96$	$L_1 = 2$ $C_1 = 96$
II	$L_2 = 2$ $C_2 = 192$	$L_2 = 2$ $C_2 = 192$
III	$L_3 = 4$ $C_3 = 384$	$L_3 = 15$ $C_3 = 384$
IV	$L_4 = 2$ $C_4 = 768$	$L_4 = 2$ $C_4 = 768$

Table 3. Training configurations and hyperparameters used for the MambaBCD-Tiny and MambaBCD-Small models

Hyperparameter	Value
Optimizer	AdamW
Learning rate	$1e^{-4}$
Weight decay	$5e^{-3}$
Batch size	Small: 8 Tiny: 16
Iterations	Small: 320000 Tiny: 640000

Details include the optimizer, learning rate, weight decay, batch sizes, number of iterations, and data augmentation techniques applied during model training.

For the datasets, we preprocess the multitemporal image pairs and their corresponding labels by cropping them to 256 × 256 pixels before inputting them into the network. We then perform inference on the original-sized data in the test set using the trained networks. Table 3 shows the training configurations, we utilize the AdamW optimizer (Loshchilov and Hutter, 2017) with a learning rate of $1e^{-4}$ and a weight decay of $5e^{-3}$, and we set the batch size to 16 for the MambaBCD-Tiny configuration and 8 for the MambaBCD-Small configuration. We conduct 640,000 training iterations for the MambaBCD-Tiny configuration and 320,000 training iterations for the MambaBCD-Small configuration both result in 40,000 epochs as the epochs are computed by $epochs = \frac{iterations}{batch\ size}$. Data augmentation techniques from PyTorch Transformers Library (*Transforming and Augmenting Images — Torchvision Main Documentation*, n.d.) such as random rotations, horizontal flips, and vertical flips are applied during training.

Given the model architectures, network frameworks, and model parameters, we experimented with the standalone MambaBCD model and our proposed Synthetic Data-augmented Mamba-Based Change Detection Algorithm (SAMBA) with its data configurations discussed in Section 3.2.

3.3. Evaluation Metrics

Precision, Recall, F1-score, Intersection over Union (IoU), overall accuracy (OA), and Kappa coefficient (KC) are critical metrics used to evaluate the performance of detection models (Kang et al., 2022; Lee et al., 2023). These metrics offer insights into the model's ability to correctly identify changes and avoid incorrect predictions. The change detection evaluation was conducted

based on the following formulas:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1\text{–}score = 2 \times \frac{Presion \times Recall}{Precision + Recall} \quad (8)$$

$$IoU = \frac{DetectionResult \cap GroundTruth}{DetectionResult \cup GroundTruth} \quad (9)$$

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$KC = \frac{P_o - P_e}{1 - P_e} \quad (11)$$

Precision measures the proportion of predicted changes that are actual changes, with true positives (TP) representing correctly identified changed pixels and false positives (FP) being unchanged pixels incorrectly classified as changed. Higher precision indicates greater reliability by minimizing false alarms. Recall evaluates the proportion of actual changes that the model correctly identifies, where false negatives (FN) are actual changes missed by the model. It complements precision by focusing on capturing all relevant changes. True negatives (TN), which are correctly identified unchanged pixels, contribute to the calculation of OA. OA assesses the model's general performance in classifying both changed and unchanged pixels but does not fully account for imbalances between the classes. The F1-score balances precision and recall, reflecting the model's ability to maintain detection accuracy while minimizing FP and FN. IoU quantifies spatial prediction accuracy by measuring the overlap between predicted and actual change regions, ensuring precise localization. The KC adjusts for random agreement, providing a robust measure of reliability by considering both observed agreement and chance. All metrics were selected based on previous change detection studies (Cao and Huang, 2023; Chen et al., 2023; Rosenfield and Fitzpatrick-Lins, 1986; Wu et al., 2019). For all these metrics, higher values are indicative of better model performance. Together, they provide a comprehensive evaluation of the model's accuracy, spatial precision, and reliability in detecting changes, ensuring it meets the demands of high-resolution change detection tasks.

4. Results

We evaluated six different configurations based on three schemes (KOMPSAT, w/ SyntheWorld (SN), and w/ SyntheWorld (BN)) and two models (MambaBCD-Tiny and MambaBCD-Small) on a change detection task using high-resolution imagery. The task involves detecting changes at the pixel level within the KOMPSAT 3,200 × 3,200 satellite image, resulting in a total of 10,240,000 pixels to classify. Each pixel is classified as either "change" (positive) or "no change" (negative). The confusion matrices summarize the performance of each model in terms of true negatives (TN), false positives (FP), false negatives (FN), and true positives (TP). The confusion matrices of these 6 configurations are displayed in Fig. 7.

Tables 4, 5, 6, and 7 summarize the performance of the MambaBCD-Small (S) and MambaBCD-Tiny (T) models across four datasets: KOMPSAT, LEVIR-CD+, WHU-CD, and SYSU-CD. Each table evaluates models trained under three configurations: KOMPSAT alone, KOMPSAT with SyntheWorld small off-nadir augmentation (SN), and KOMPSAT with SyntheWorld big off-nadir augmentation (BN). Key metrics include F1-score, Precision,

Table 4. Validation performance comparison of different MambaBCD model configurations on the KOMPSAT dataset

Training dataset	KOMPSAT		w/ SyntheWorld (SN)		w/ SyntheWorld (BN)	
Mamba model	S	T	S	T	S	T
F1 score	96.622	**97.434**	95.477	**96.865**	96.390	96.736
Precision	97.953	**98.340**	97.253	**98.024**	97.519	97.814
Recall	95.327	**96.544**	93.764	**95.733**	95.287	95.681
IoU	93.466	**94.997**	91.345	**93.921**	93.032	93.678
Kappa coefficient	96.436	**97.292**	95.228	**96.692**	96.191	96.555
Overall accuracy	99.647	**99.730**	99.529	**99.671**	99.621	99.657

The table evaluates the Small (S) and Tiny (T) configurations of the MambaBCD model trained on three datasets: KOMPSAT alone, KOMPSAT with SyntheWorld using small off-nadir images (SN), and KOMPSAT with SyntheWorld using big off-nadir images (BN). Key performance metrics include F1-score, precision, recall, Intersection over Union (IoU), Kappa coefficient, and overall accuracy. The table highlights the highest metric values in **red** and the second-highest values in **blue**, showcasing the top-performing configurations.

https://doi.org/10.7780/kjrs.2024.40.6.3.11

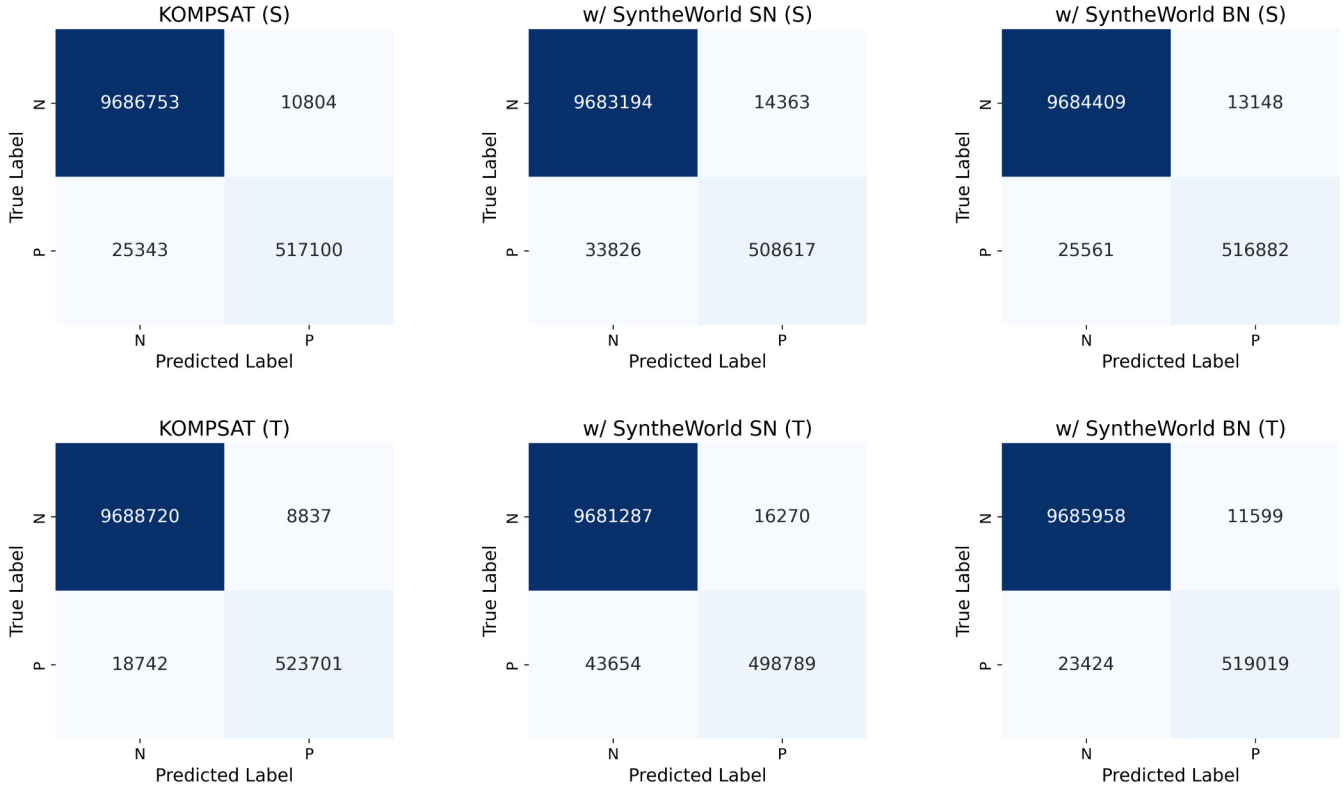

Fig. 7. Confusion matrices for various configurations of the MambaBCD model (KOMPSAT, SN, and BN versions in MambaBCD-Small and MambaBCD-Tiny architectures) are presented here, each labeled with its respective epoch number. The matrices illustrate the model's performance on Binary Change Detection, showing the counts of True Negatives (TN), False Positives (FP), False Negatives (FN), and True Positives (TP). The primary diagonal represents accurate predictions, with higher values indicating better performance in correctly identifying changes and non-changes across different configurations.

Recall, IoU, KC, and OA. Across all datasets, the MambaBCD-Tiny model demonstrates its effectiveness, particularly when paired with appropriate training data configurations.

Table 4 highlights the superior performance of the MambaBCD-Tiny model trained solely on KOMPSAT data, achieving the highest metrics across all categories, including F1-score (97.434%), precision (98.340%), recall (96.544%), IoU (94.997%), KC (97.292%), and OA (99.730%). These results demonstrate the effectiveness of the Tiny model in capturing detailed spatial and temporal changes in high-resolution satellite imagery. While the incorporation of

Table 5. Validation performance comparison of MambaBCD model configurations on the LEVIR-CD+ dataset

Training dataset	KOMPSAT		w/ SyntheWorld (SN)		w/ SyntheWorld (BN)	
Mamba model	S	T	S	T	S	T
F1 score	25.079	23.340	20.125	**37.663**	26.791	**30.087**
Precision	35.024	**38.581**	34.834	**42.526**	38.432	36.346
Recall	19.533	16.730	14.145	**33.798**	20.562	**25.667**
IoU	14.337	13.212	11.188	**23.201**	15.467	**17.707**
Kappa coefficient	22.758	21.343	18.140	**35.257**	24.584	**27.571**
Overall accuracy	95.102	**95.388**	95.286	**95.305**	95.284	94.994

The table evaluates the Small (S) and Tiny (T) configurations of the MambaBCD model trained under three conditions: KOMPSAT alone, KOMPSAT with SyntheWorld using small off-nadir images (SN), and KOMPSAT with SyntheWorld using big off-nadir images (BN). Performance metrics include F1-score, precision, recall, Intersection over Union (IoU), Kappa coefficient, and overall accuracy. The highest values for each metric are highlighted in **red**, and the second-highest values are highlighted in **blue**, illustrating the effectiveness of various training configurations. Notably, the Tiny model trained with SyntheWorld (SN) demonstrates superior performance in several metrics, underscoring the benefits of data augmentation in specific contexts.

Table 6. Validation performance comparison of MambaBCD model configurations on the WHU-CD dataset

Training dataset	KOMPSAT		w/ SyntheWorld (SN)		w/ SyntheWorld (BN)	
Mamba model	S	T	S	T	S	T
F1 score	44.000	**49.375**	44.707	**48.315**	47.223	46.246
Precision	32.640	37.533	**37.960**	36.497	36.316	**37.546**
Recall	67.490	**72.133**	54.370	**71.449**	67.496	60.195
IoU	28.205	**32.780**	28.789	**31.852**	30.910	30.078
Kappa coefficient	39.417	**45.340**	40.794	**44.167**	43.069	42.255
Overall accuracy	90.360	91.699	**92.453**	91.421	91.534	**92.147**

The table evaluates the Small (S) and Tiny (T) versions of the MambaBCD model trained under three conditions: KOMPSAT alone, KOMPSAT with SyntheWorld using small off-nadir images (SN), and KOMPSAT with SyntheWorld using big off-nadir images (BN). Performance metrics include F1-score, precision, recall, Intersection over Union (IoU), Kappa coefficient, and overall accuracy. The highest values in each metric are highlighted in **red**, while the second highest are marked in **blue**, showcasing the relative performance of different training configurations. The Tiny model trained with KOMPSAT alone achieved the highest F1-score, recall, and IoU, demonstrating its ability to effectively capture changes in this dataset.

SyntheWorld data (SN and BN) improved some metrics for the MambaBCD-Small model, the KOMPSAT-trained Tiny model consistently outperformed others.

The validation results on the LEVIR-CD+ dataset highlight the impact of synthetic data augmentation (Table 5). The MambaBCD-Tiny model trained with SyntheWorld SN data outperformed other configurations, achieving the highest F1-score (37.663%), Precision (42.526%), Recall (33.798%), IoU (23.201%), and KC (35.257%). This demonstrates that SyntheWorld small off-nadir data effectively enhances the model's ability to generalize to datasets with diverse structural changes. The KOMPSAT-trained models performed reasonably well but were outpaced by configurations that incorporated synthetic data, particularly for the Tiny model, emphasizing the value of tailored data augmentation strategies.

The validation results for the WHU-CD dataset reveal that the MambaBCD-Tiny model trained on KOMPSAT data alone achieved the highest F1-score (49.375%), Recall (72.133%), IoU (32.780%), and KC (45.340%) (Table 6). These findings highlight the Tiny model's robustness in detecting changes in high-resolution urban imagery without synthetic data. Meanwhile, the MambaBCD-Small model trained with SyntheWorld SN data delivered competitive results, particularly for recall (54.370%) and precision (37.960%), demonstrating that synthetic data can benefit certain metrics, especially for smaller models.

On the SYSU-CD dataset (Table 7), where the MambaBCD-Tiny model trained on KOMPSAT data achieved the highest F1-score (16.215%), IoU (8.823%), KC (12.312%), and OA (79.882%). Despite this, the recall values for all configurations remained low, indicating challenges in detecting all true changes in this dataset. Precision was consistently high across all models, with the best precision (92.113%) observed in the Tiny model trained with SyntheWorld BN data. However, synthetic data augmentation did not significantly enhance overall performance, suggesting that the KOMPSAT dataset alone is better suited for training models

Table 7. Validation performance comparison of MambaBCD model configurations on the SYSU-CD dataset

Training dataset	KOMPSAT		w/ SyntheWorld (SN)		w/ SyntheWorld (BN)	
Mamba model	S	T	S	T	S	T
F1 score	**10.002**	**16.215**	7.647	9.232	6.365	4.681
Precision	81.618	78.417	84.558	**88.855**	85.418	**92.113**
Recall	**5.327**	**9.043**	4.000	4.869	3.305	2.402
IoU	**5.264**	**8.823**	3.975	4.839	3.287	2.397
Kappa coefficient	7.563	**12.312**	5.813	7.155	4.838	3.627
Overall accuracy	79.359	**79.882**	79.175	**79.387**	79.061	78.943

The table reports results for the Small (S) and Tiny (T) versions of the MambaBCD model trained on three datasets: KOMPSAT alone, KOMPSAT with SyntheWorld using small off-nadir images (SN), and KOMPSAT with SyntheWorld using big off-nadir images (BN). Metrics include F1-score, precision, recall, Intersection over Union (IoU), Kappa coefficient, and overall accuracy. The highest values in each metric are highlighted in **red**, and the second-highest values are in **blue**, emphasizing the best-performing configurations. Notably, while precision is high across all configurations, recall, and IoU remain low, suggesting challenges in detecting all true changes, particularly with SyntheWorld-augmented datasets

https://doi.org/10.7780/kjrs.2024.40.6.3.11

on SYSU-CD.

The results in the table collectively demonstrate that the MambaBCD-Tiny model consistently outperforms its larger counterpart across most datasets, often achieving superior results despite having fewer parameters. Synthetic data augmentation through SyntheWorld SN and BN improves performance in certain datasets, notably LEVIR-CD+ and WHU-CD, while its impact is less pronounced for KOMPSAT and SYSU-CD. These results underscore the importance of adapting training configurations and data augmentation strategies to the unique characteristics of each dataset to maximize model effectiveness.

The models demonstrated strong performance in several change categories (Fig. 8). Specifically, the models performed favorably for categories (2) building demolition, (4) and (5) on-going construction, and (6) no change, consistently detecting changes or confirming unchanged areas accurately. In category (3) Building renovation/replacement, while the models generally performed well, some false negatives were observed, particularly with very small buildings, indicating challenges in detecting subtle changes in smaller structures.

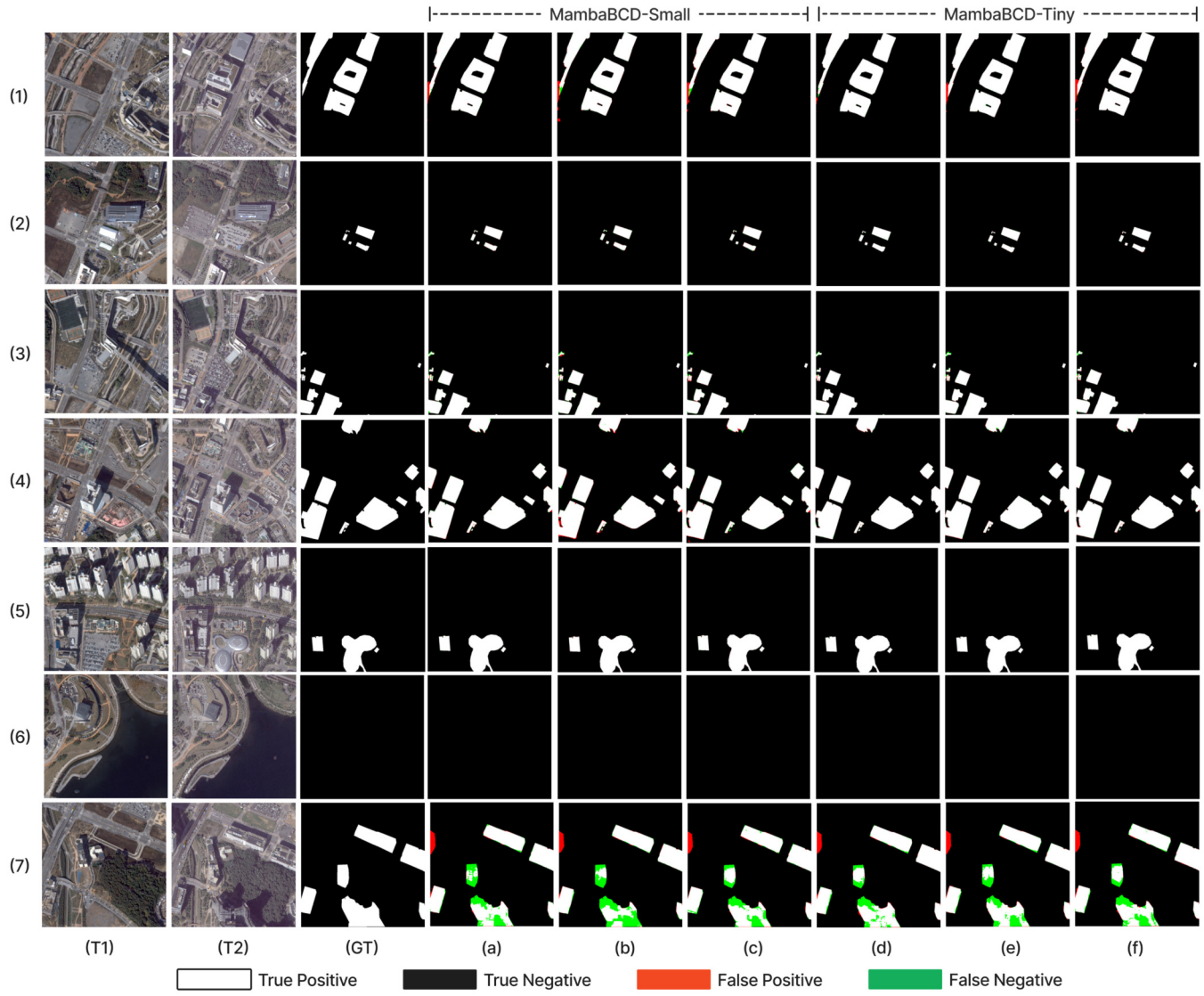

Fig. 8. Change maps across different conditions: (1) Building construction (2) Building demolition (3) Building renovation/replacement (4-5) Building on-going construction (6) No change (7) Result insights. Based on the results of the different model types, and with and without the SyntheWorld small off-nadir (SN) and big off-nadir (BN) data configurations: (a) KOMPSAT MambaBCD-Small (b) KOMPSAT+SyntheWorld (SN) MambaBCD-Small (c) KOMPSAT+SyntheWorld (BN) MambaBCD-Small (d) KOMPSAT MambaBCD-Tiny (e) KOMPSAT+SyntheWorld (SN) MambaBCD-Tiny (f) KOMPSAT+SyntheWorld (BN) MambaBCD-Tiny. (T1) refers to the pre-event images, (T2) refers to the post-event images, and (GT) is the ground truth binary change map.

For category (1) building construction and the insights derived from category (7), the models exhibited false positives. However, further inspection of the input data revealed that some changes were not properly marked in the ground truth data, suggesting potential annotation gaps. Despite these omissions in the ground truth, it is notable that the models were still able to detect such changes, showcasing their sensitivity and capability to recognize alterations even when not explicitly indicated. Additionally, in category (7), false negatives were noted, which appeared to stem from inaccuracies in the ground truth data where the building shapes were not ideally masked. This inconsistency likely led to some confusion during the model's inference, affecting the output's accuracy. These observations underscore the models' robust performance while also highlighting the importance of precise ground truth data for optimal change detection outcomes.

5. Discussion

5.1. Performance Evaluation of Mamba Models

The robust accuracy of the MambaBCD models, particularly the Tiny configuration, can be attributed to the architecture's efficient handling of spatiotemporal relationships inherent in VHR satellite imagery. The Tiny model, with its streamlined architecture featuring fewer VSS blocks and channels, demonstrated superior generalization capabilities compared to the Small variant. This efficiency likely reduces the risk of overfitting, allowing the Tiny model to maintain high performance even with limited training data. Additionally, the Siamese encoder structure ensures consistent feature extraction from multitemporal image pairs, enhancing the model's ability to accurately discern changes by maintaining uniformity in feature representation across different time points. This architectural design, as detailed by Gu and Dao (2024), proved especially effective on datasets like LEVIR-CD+, which share structural and temporal characteristics with the KOMPSAT-3A imagery, facilitating seamless integration and high-precision change detection.

The exceptional performance of the MambaBCD-Tiny model on datasets such as LEVIR-CD+ and WHU-CD can be further explained by the alignment between these benchmark datasets and the training data derived from KOMPSAT-3A imagery. LEVIR-CD+, constructed using a combination of Google Earth and satellite datasets, closely mirrors the high-resolution and multitemporal aspects of KOMPSAT-3A images. This similarity enables the MambaBCD-Tiny model to effectively leverage its SSM foundation, capturing intricate structural details and temporal dynamics essential for accurate change detection (Chen et al., 2024a). In contrast, the lower performance on the SYSU-CD dataset highlights the model's dependency on dataset-specific characteristics. The SYSU-CD dataset, differing in geographical context and urban complexity, presents challenges that the current MambaBCD architecture was not specifically optimized to address, emphasizing the importance of dataset alignment for achieving optimal model performance (Paranjape et al., 2024).

Overall, MambaBCD's architecture provides a robust framework for high accuracy change detection in datasets that share structural and temporal similarities with the training data. Its ability to efficiently process and model complex spatiotemporal information makes it particularly well-suited for applications involving high-resolution satellite imagery, such as KOMPSAT-3A and LEVIR-CD+ (Gu and Dao, 2024). However, the varying performance across different datasets emphasizes the need for careful consideration of dataset characteristics and potentially tailored architectural adjustments to maintain robustness and accuracy in more heterogeneous remote sensing environments.

5.2 Impact of Synthetic Dataset in Change Detection

The incorporation of synthetic data from SyntheWorld significantly enhanced the performance of the MambaBCD models, particularly on the LEVIR-CD+ benchmark dataset. Our analysis indicates that the diverse range of change scenarios introduced by SyntheWorld enabled the models to better generalize across varied urban landscapes and construction dynamics (Song et al., 2024). The SSM foundation of MambaBCD, combined with its selective state mechanism, allowed the architecture to effectively integrate the synthetic variations, capturing intricate spatiotemporal relationships that are essential for accurate change detection (Chen et al., 2024a). This synergy between synthetic augmentation and the robust architectural design of MambaBCD facilitated higher precision and recall rates, demonstrating the model's ability to adapt to complex and diverse change patterns that may not be sufficiently represented in the limited real-world training data (Paranjape et al., 2024).

However, the benefits of synthetic data were not uniformly observed across all datasets, as evidenced by the limited improvements on the SYSU-CD dataset. This discrepancy suggests that the effectiveness of synthetic augmentation is contingent upon the alignment between the synthetic data characteristics and those of the target dataset. In cases where the synthetic data closely mirrors the real-world variations present in the target dataset, as with LEVIR-CD+, the MambaBCD

https://doi.org/10.7780/kjrs.2024.40.6.3.11

models leveraged their architectural strengths to achieve robust accuracy (Song et al., 2020). Conversely, for datasets with distinct geographical or urban complexities not adequately captured by the synthetic scenarios, the models showed less improvement (Chen et al., 2021). Our analysis underscores the importance of tailoring synthetic data generation to match the specific attributes of each application context. Ultimately, the strategic integration of synthetic datasets can significantly expand the applicability and reliability of change detection models like MambaBCD, enabling more accurate and resilient monitoring across diverse remote sensing environments (Zhang et al., 2022).

6. Conclusions

This study has investigated the effectiveness of MambaBCD models for remote sensing change detection tasks, with a focus on high-resolution image analysis. By evaluating multiple configurations of the MambaBCD architecture, including Original, Spectral Normalization, and Batch Normalization models in both Small and Tiny versions, we have demonstrated the adaptability and performance potential of this architecture across diverse datasets. The results indicate that the MambaBCD-Tiny model consistently delivers competitive results, combining high accuracy in detecting changes with computational efficiency—a key consideration in practical applications requiring scalability.

Our findings suggest that MambaBCD models excel at distinguishing large-scale changes, such as building demolitions or new construction, while presenting challenges in detecting subtler modifications. This limitation highlights areas for future improvement, particularly in developing finer-tuned models or leveraging higher-resolution datasets to capture intricate changes in smaller structures or during renovations.

While synthetic data augmentation (SyntheWorld) has shown benefits in certain datasets, such as LEVIR, its impact is inconsistent, underscoring the importance of a tailored data strategy. This study reveals that the utility of synthetic data varies depending on dataset characteristics, suggesting that data augmentation approaches should be carefully chosen to match the specifics of each application domain.

However, this study has some limitations. First, the variability in performance across different datasets points to the sensitivity of MambaBCD models to data quality and ground truth accuracy. Some false positives and false negatives were attributed to incomplete or inaccurately labeled data, indicating that high-quality, precise labeling is essential for optimal model performance. Additionally, our reliance on small-scale and medium-scale configurations due to computational constraints means that further exploration with larger architectures could provide insights into maximizing the model's capacity.

In future work, we plan to investigate more advanced data augmentation techniques, including multimodal data integration and synthetic augmentation tailored to specific dataset properties, to improve the detection of nuanced changes. We also aim to explore the potential of MambaBCD for broader remote sensing tasks, applying it to larger datasets and refining its architecture for enhanced performance in complex, real-world change detection scenarios.

Acknowledgments

This research was supported by data provided by the Korea Aerospace Research Institute (KARI) through the 2024 Satellite Information Utilization Competition, and by a grant (RS-2022-00155763) from the Development of Comprehensive Land Management Technology Using Satellite Image Information Big Data Project funded by the Ministry of Land, Infrastructure, and Transport of the Korean government.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

Cao, Y., and Huang, X., 2023. A full-level fused cross-task transfer learning method for building change detection using noise-robust pretrained networks on crowdsourced labels. *Remote Sensing of Environment, 284, 113371*. https://doi.org/10.1016/j.rse.2022.113371

Chen, H., and Shi, Z., 2020. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing, 12*(10), 1662. https://doi.org/10.3390/rs12101662

Chen, H., Song, J., Han, C., Xia, J., and Yokoya, N., 2024a. ChangeMamba: Remote sensing change detection with spatiotemporal state space model. *arXiv preprint arXiv: 2404.03425*. https://doi.org/10.48550/arXiv.2404.03425

Chen, H., Song, J., Wu, C., Du, B., and Yokoya, N., 2023. Exchange means change: An unsupervised single-temporal change

detection framework based on intra- and inter-image patch exchange. *ISPRS Journal of Photogrammetry and Remote Sensing, 206*, 87–105. https://doi.org/10.1016/j.isprsjprs.2023.11.004

Chen, H., Qi, Z., and Shi, Z., 2021. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–14. https://doi.org/10.1109/TGRS.2021.3095166

Chen, K., Chen, B., Liu, C., Li, W., Zou, Z., and Shi, Z., 2024b. RSMamba: Remote sensing image classification with state space model. *IEEE Geoscience and Remote Sensing Letters, 21*, 8002605. https://doi.org/10.1109/LGRS.2024.3407111

Cheng, G., Huang, Y., Li, X., Lyu, S., Xu, Z., Zhao, H., et al., 2024. Change detection methods for remote sensing in the last decade: A comprehensive review. *Remote Sensing, 16*(13), 2355. https://doi.org/10.3390/rs16132355

Daudt, R. C., Le Saux, B., and Boulch, A., 2018. Fully convolutional Siamese networks for change detection. In *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, Oct. 7–10, pp. 4063–4067. https://doi.org/10.1109/ICIP.2018.8451652

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. https://doi.org/10.48550/arXiv.2010.11929

Gu, A., and Dao, T., 2024. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*. https://doi.org/10.48550/arXiv.2312.00752

Han, Y., Kim, T., Han, S., and Song, J., 2017. Change detection of urban development over large area using KOMPSAT optical imagery. *Korean Journal of Remote Sensing, 33*(6–3), 1223–1232. https://doi.org/10.7780/kjrs.2017.33.6.3.6

Hussain, M., Chen, D., Cheng, A., Wei, H., and Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing, 80*, 91–106. https://doi.org/10.1016/j.isprsjprs.2013.03.006

Im, J., Jensen, J. R., and Tullis, J. A., 2008. Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing, 29*(2), 399–423. https://doi.org/10.1080/01431160601075582

Jeon, M.-J., Lee, S.-R., Kim, E., Lim, S.-B., and Choi, S.-W., 2016. Launch and early operation results of KOMPSAT-3A. In *Proceedings of the 14th International Conference on Space Operations*, Daejeon, Korea, May 16–20, pp. 1–10. https://doi.org/10.2514/6.2016-2394

Kang, Y., Jang, E., Im, J., and Kwon, C., 2022. A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. *GIScience & Remote Sensing, 59*(1), 2019–2035. https://doi.org/10.1080/15481603.2022.2143872

Kim, N., Choi, Y., Bae, J., and Sohn, H. G., 2020. Estimation and improvement in the geolocation accuracy of rational polynomial coefficients with minimum GCPs using KOMPSAT-3A. *GIScience & Remote Sensing, 57*(6), 719–734. https://doi.org/10.1080/15481603.2020.1791499

Lee, C., Yun, Y., Bae, S., Eo, Y. D., Kim, C., Shin, S., et al., 2021. Analysis of deep learning research trends applied to remote sensing through paper review of Korean domestic journals. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 39*(6), 437–456. https://doi.org/10.7848/ksgpc.2021.39.6.437

Lee, D., Kim, J., and Kim, Y., 2024. Optimal hyperparameter analysis of segment anything model for building extraction using KOMPSAT-3/3A images. *Korean Journal of Remote Sensing, 40*(5–1), 551–568. https://doi.org/10.7780/kjrs.2024.40.5.1.11

Lee, S., Kang, Y., Sung, T., and Im, J., 2023. Efficient deep learning approaches for active fire detection using Himawari-8 geostationary satellite images. *Korean Journal of Remote Sensing, 39*(5–3), 979–995. https://doi.org/10.7780/kjrs.2023.39.5.3.8

Liu, J., and Ji, S., 2020. A novel recurrent encoder-decoder structure for large-scale multi-view stereo reconstruction from an open aerial dataset. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 13–19, pp. 6050–6059. https://doi.org/10.1109/CVPR42600.2020.00609

Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., et al., 2024. VMamba: Visual state space model. *arXiv preprint arXiv:2401.10166*. https://doi.org/10.48550/arXiv.2401.10166

Loshchilov, I., and Hutter, F., 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*. https://arxiv.org/abs/1711.05101

Lu, K., Huang, X., Xia, R., Zhang, P., and Shen, J., 2024. Cross attention is all you need: Relational remote sensing change detection with transformer. *GIScience & Remote Sensing, 61*(1), 2380126. https://doi.org/10.1080/15481603.2024.

2380126

Paranjape, J. N., de Melo, C., and Patel, V. M., 2024. A Mamba-based Siamese network for remote sensing change detection. *arXiv preprint arXiv:2407.06839*. https://doi.org/10.48550/arXiv.2407.06839

Park, S., and Song, A., 2023. Hybrid approach using deep learning and graph comparison for building change detection. *GIScience & Remote Sensing, 60*(1), 2220525. https://doi.org/10.1080/15481603.2023.2220525

Peng, D., Zhang, Y., and Guan, H., 2019. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sensing, 11*(11), 1382. https://doi.org/10.3390/rs11111382

Rosenfield, G. H., and Fitzpatrick-Lins, K., 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing, 52*(2), 223–227.

Shi, Q., Liu, M., Li, S., Liu, X., Wang, F., and Zhang, L., 2022. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–16. https://doi.org/10.1109/TGRS.2021.3085870

Song, C., Wahyu, W., Jung, J., Hong, S., Kim, D., and Kang, J., 2020. Urban change detection for high-resolution satellite images using U-Net based on SPADE. *Korean Journal of Remote Sensing, 36*(6–2), 1579–1590. https://doi.org/10.7780/kjrs.2020.36.6.2.8

Song, J., Chen, H., and Yokoya, N., 2024. SyntheWorld: A large-scale synthetic dataset for land cover mapping and building change detection. In *Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, Jan. 3–8, pp. 8272–8281. https://doi.org/10.1109/WACV57701.2024.00810

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al., 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*. https://doi.org/10.48550/arXiv.1706.03762

Wu, C., Chen, H., Do, B., and Zhang, L., 2019. Unsupervised change detection in multi-temporal VHR images based on deep kernel PCA convolutional mapping network. *IEEE Transactions on Cybernetics, 52*(11), 12084–12098. https://doi.org/10.1109/TCYB.2021.3086884

Zhang, C., Wang, L., Cheng, S., and Li, Y., 2022. SwinSUNet: Pure transformer network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–13. https://doi.org/10.1109/TGRS.2022.3160007

Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., and Wang, X., 2024. Vision Mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*. https://doi.org/10.48550/arXiv.2401.09417